CLAIMS

<u>What is claimed is:</u>

Sub A4

1. A method of populating a data structure with a plurality of character strings, said method comprising:

i) encoding two or more a biological molecules into character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a collection of strings; and

v) optionally repeating steps (i) or (ii) through (iv) using one or more of said product strings as an initial string in the collection of initial character strings.

2. The method of claim 1, wherein said encoding comprises encoding one or more nucleic acid sequences into said character strings.

Sub C1

3. The method of claim 2, wherein said one or more nucleic acid sequences comprise a nucleic acid sequence encoding a known protein.

4. The method of claim 1, wherein said encoding comprises encoding one or more amino acid sequences into said character strings.

5. The method of claim 4, wherein said one or more amino acid sequences comprise a nucleic acid sequence encoding a known protein.

Sub A5

6. The method of claim 1, wherein said biological molecules have at least 30% sequence identity.

Sub E3

7. The method of claim 1, wherein said selecting comprises selecting substrings such that the ends of said substrings occur in string regions of about 3 to about 20 characters that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the same two strings.

8. The method of claim 1, wherein said selecting comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

9. The method of claim 1, wherein said selecting and concatenating comprises concatenating substrings from two different initial strings such that the concatenation occurs in a region of about three to about twenty characters having higher sequence identity between said two different initial strings than the overall sequence identity between said two different initial strings.

Sub B3

10. The method of claim 1, wherein said selecting comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the character strings, and selecting a character that is a member of an aligned pair for the end of one substring.

Sub C3

11. The method of claim 1, wherein said product strings are added to the collection only if they have greater than 30% sequence identity with the initial strings.

12. The method of claim 1, wherein said method further comprises randomly altering one or more characters of said character strings.

13. The method of claim 12, wherein said method further comprises randomly selecting and altering one or more occurrences of a particular preselected character in said character strings.

14. The method of claim 1, wherein said coding, selecting, or concatenating is performed on an internet site.

[illegible]

15. The method of claim 1, wherein said coding, selecting, or concatenating is performed on a server. The method of claim 1, wherein said coding, selecting, or concatenating is performed on a server.

16. The method of claim 1, wherein said coding, selecting, or concatenating is performed on a server. The method of claim 1, wherein said coding, selecting, or concatenating is performed on a client linked to a network..

~~17.~~ A computer program product comprising computer code that

i) encodes two or more a biological molecules into character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;

Sub C3

ii) selects at least two substrings from said character strings;

iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adds the product strings to a collection of strings; and

v) optionally repeats steps (i) or (ii) through (iv) using one or more of said product strings as an initial string in the collection of initial character strings.

18. The program of claim 17, wherein said two or more biological molecules are nucleic acid sequences.

19. The program of claim 17, wherein said two or more biological molecules are nucleic acid sequences of known proteins.

20. The program of claim 17, wherein said two or more biological molecules are amino acid sequences

21. The program of claim 17, wherein said biological molecules have at least 30% sequence identity.

22. The program of claim 17, wherein said code selects substrings such that the ends of said substrings occur in string regions of about three to about twenty characters that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the same two strings.

23. The program of claim 17, wherein said code selects substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

24. The program of claim 17, wherein said code selects and concatenates substrings from two different initial strings such that the concatenation occurs in a region of about three to about twenty characters having higher sequence identity between said two

different initial strings than the overall sequence identity between said two different initial strings.

Sub C2

25. The program of claim 17, wherein code selects substrings by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the character strings, and selecting a character that is a member of an aligned pair for the end of one substring.

C2 cont

26. The program of claim 17, wherein said product strings are added to the collection only if they have greater than 30% identity with the initial strings.

[illegible]

27. The program of claim 17, wherein said method further comprises randomly altering one or more characters of said character strings.

28. The program of claim 27, wherein said method further comprises randomly selecting and altering one or more occurrences of a particular preselected character in said character strings.

Sub C3

29. The program claim 17, wherein said code is stored on media selected from the group consisting of magnetic media, optical media, optomagnetic media.

30. The program claim 17, wherein said code is in dynamic or static memory of a computer.

31. A label generating system for creating a plurality of related labels, said labeling system comprising:

an encoder for encoding two or more initial strings from biological molecules;

an isolator for identifying and selecting substrings from said two or more strings;

a concatenator for concatenating said substrings;

a data structure for storing the concatenated substrings as a collection of strings;

a comparator for measuring the number and variability of the collection of strings and determining that sufficient strings exist in the collection of strings; and

a command writer for writing the collection of strings into a raw string file.

32. The system of 31, wherein said isolator comprises a comparator for aligning and determining regions of identity between said two or more initial strings;

33. The system of 31, wherein said encoder comprises a means for encoding a nucleic acid sequence into a character string.

34. The system of 31, wherein said encoder comprises a means for encoding an amino acid sequence into a character string.

35. The system of claim 31, wherein said comparator comprises a means for calculating sequence identity.

36. The system of claim 31, wherein said isolator selects substrings such that the ends of said substrings occur in string regions of about three to about 100 characters that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the same two strings.

37. The system of claim 31, wherein said isolator selects substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

38. The system of claim 31, wherein said isolator and concatenator individually or in combination concatenate substrings from two different initial strings such that the concatenation occurs in a region of about three to about 100 characters having higher sequence identity between said two different initial strings than the overall sequence identity between said two different initial strings.

39. The system of claim 31, wherein said isolator aligns two or more of said initial character strings to maximize pairwise identity between two or more substrings of the character strings, and selecting a character that is a member of an aligned pair for the end of one substring.

40. The system of claim 31, wherein said comparator adds strings to said data structure only if they have greater than 30% identity with the initial strings.

41. The system of claim 31, further comprising an operator to randomly altering one or more characters of the character strings.

42. The system of claim 41, wherein said operator randomly selects and alters one or more occurrences of a particular ~~preselected~~ character in said character strings.

43. The system of claim 31, wherein data structure is a data structure that stores encoded nucleic acid sequences.

44. The system of claim 31, wherein data structure is a data structure that stores encoded amino acid sequences.

ADD C3 ADD E5